Probability and Statistics for JMC

Lectured by Alex Geringer-Sameth

Typed by Aris Zhu Yi Qing

May 15, 2021

Contents

Chapter 1

Review of Elementary Set Theory

Ω	universal set
$\emptyset$	empty set
$A \subseteq \Omega$	subset of Ω
$\overline{A}$	Complement of A
$\lvert A \rvert$	cardinality of A
$A \cup B$	union (A or B)
$A \cap B$	intersection(A and B)
$A = B$	both sets have exactly the same elements
$A \backslash B$	set difference (elements in A that are not in B)
$\{\omega\}$	a singleton with only the element ω in the set
$A \times B$	$\big\{(a,b) \lvert a \in A, b \in B\big\}$

Chapter 2

Visual and Numerical Summaries

2.1 Visualization

Definition 1. The ***histogram*** allows us to visualize how a sample of data is distributed, say the observed values are $\{x_1, \ldots, x_2\}$. The first step is deciding on a set of ***bins*** that divide the range of x into a series of intervals. A histogram then shows the ***frequency*** for each bin.

Comments Often the histogram's y-axis is normalized in some way.

- Instead of showing frequency, the height of the histogram can show ***relative frequency***, the fraction of the data set contained within the bin. In this case, $1 = \sum_{\text{bins } i} y_i$, where y_i is the relative frequency at bin i.
- The histogram could also show the ***density***, the relative frequency divided by the bin width. In this case, $1 = \sum_{\text{bins } i} \rho_i \Delta x_i$, where ρ_i is the density for bin i and Δx_i is the width of bin i.

Definition 2. The ***empirical cumulative distribution function*** of a sample of real values $\{x_1, \ldots, x_n\}$ is

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(x_i \leq x),$$

where $I(x_i \leq x)$ is an ***indicator function***, i.e. the value is 1 when $x_i \leq x$ and 0 when $x_i > x$.

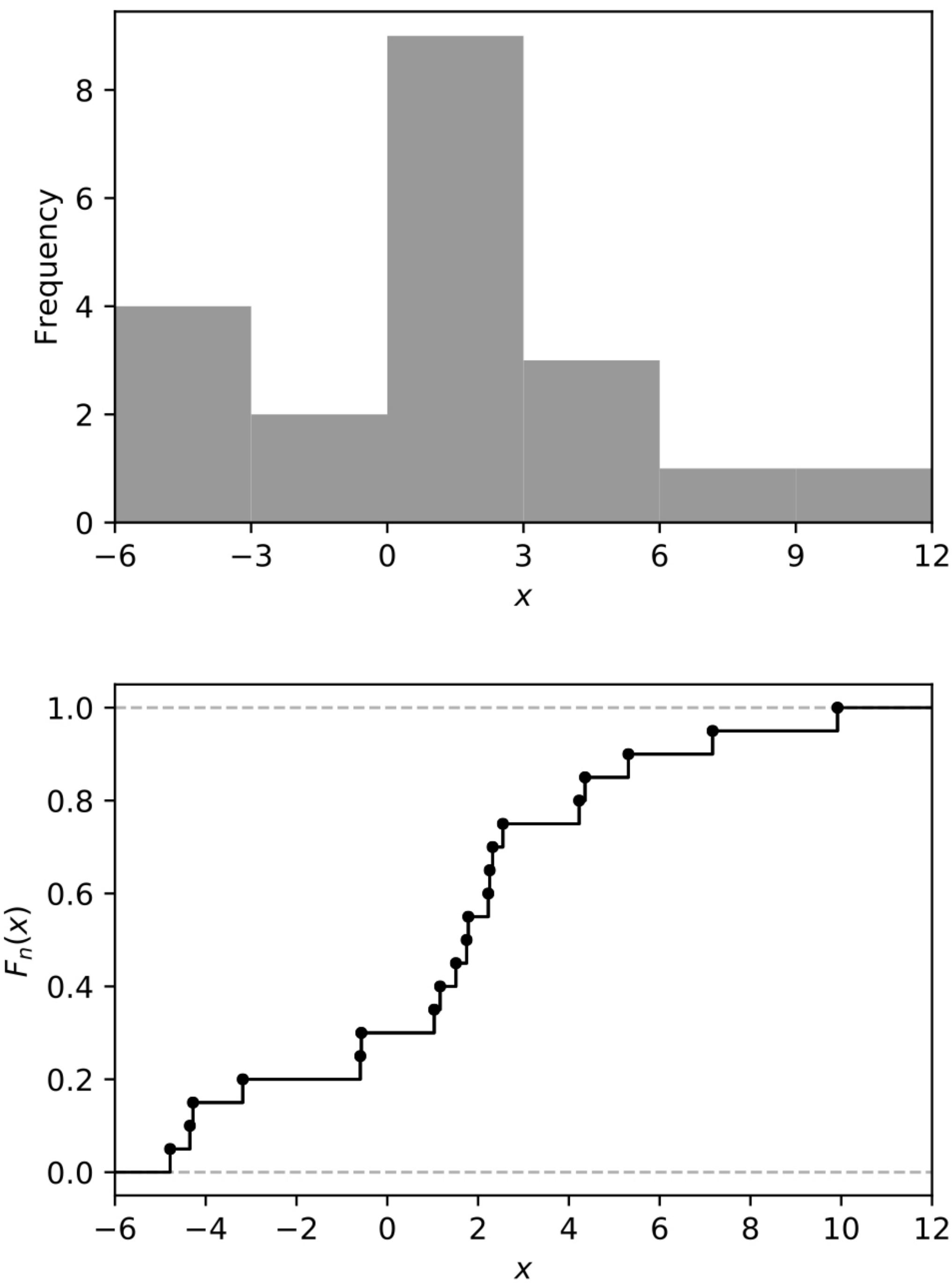

Figure 2.1: The first diagram is the histogram, and the second diagram is the empirical cdf with the same set of data

2.2 Summary Statistics

2.2.1 Measures of Location

arithmetic mean	$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
geometric mean	$x_G = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$
harmonic mean	$\frac{1}{x_H} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{x_i}$
i^{th} order statistic	$x_{(i)}$ = the i^{th} smallest value of the sample
median	$x_{\left(\frac{n+1}{2}\right)}$
mode	x_i which occurs most frequently in the sample

Comments

- For positive data $\{x_1, \ldots, x_n\}$,

$$\text{arithmetic mean} \geq \text{geometric mean} \geq \text{harmonic mean.}$$

- Arithmetic mean and geometric mean are related in the following way:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n} \ln y_i = \frac{1}{n}\ln\prod_{i=1}^{n} y_i = \ln\left(\prod_{i=1}^{n} y_i\right)^{\frac{1}{n}} = \ln x_G,$$

  where $x_i = \ln y_i$.

- For $x_{(i)}$, when i is not an integer, we define $\alpha \in (0,1)$ s.t. $\alpha = i - \lfloor i \rfloor$, and

$$x_{(i)} = (1-\alpha)x_{(\lfloor i \rfloor)} + \alpha x_{(\lceil i \rceil)}.$$

2.2.2 Measures of Dispersion

mean square/sample variance	$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$
root mean square/sample standard deviation	$s = \sqrt{s^2}$
range	$x_{(n)} - x_{(1)}$
first quartile	$x_{\left(\frac{1}{4}(n+1)\right)}$
third quartile	$x_{\left(\frac{3}{4}(n+1)\right)}$
interquartile range	$x_{\left(\frac{1}{4}(n+1)\right)} - x_{\left(\frac{3}{4}(n+1)\right)}$

Comments

- sample variance's different expression:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2 = \overline{x^2} - \overline{x}^2.$$

- Robustness, shown in table 2.1

Table 2.1: Robustness of different location and dispersion statistic

	Least Robust	More Robust	Most Robust
Location	$\frac{x_{(1)}+x_{(n)}}{2}$	$\overline{x}$	$x_{\left(\frac{n+1}{2}\right)}$
Dispersion	$x_{(n)} - x_{(1)}$	s	$x_{\left(\frac{3}{4}(n+1)\right)} - x_{\left(\frac{1}{4}(n+1)\right)}$

2.2.3 Covariance, Correlation, and Skewness

covariance	$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$
correlation	$r_{xy} = \frac{s_{xy}}{s_x s_y}$
skewness	$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{s}\right)^3$

Comments

- covariance's different expression:

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i + \frac{1}{n}\sum_{i=1}^{n} -x_i\overline{y} - \overline{x}y_i + \overline{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n} - \overline{x}\,\overline{y}.$$

  In the random variable's context, it is

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

- Correlation gives a **scale-invariant** measurement of relatedness between x and y, since

$$|r_{xy}| \leq 1.$$

- A sample is **positively (negatively)** or **right (left) skewed** if the upper tail of the histogram of the sample is longer (shorter) than the lower tail.

2.2.4 Box-and-whisker plot

The diagram is based on the five-point summary (use Figure 2.2 as reference):

- Median – middle line in the box.
- 3rd and 1st Quartiles – top and bottom of the box.
- "Whiskers" – extend out as dashed lines from the box to max/min values, which are the two short horizontal lines.
- Any outliers, i.e. extreme points beyond the whiskers, are plotted individually as dots.

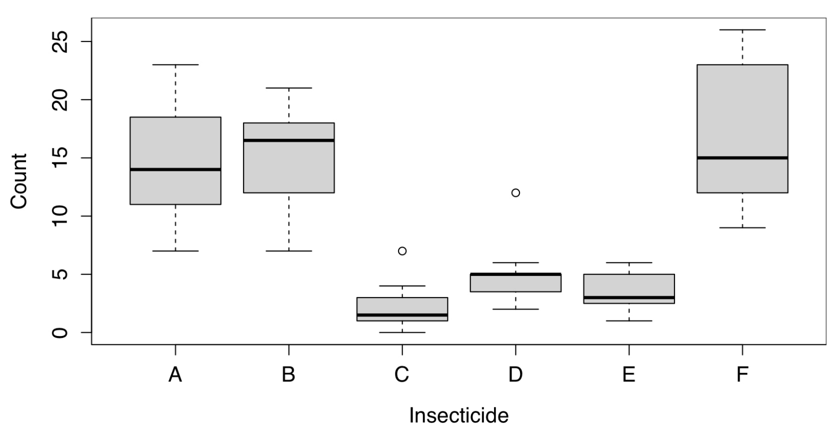

Figure 2.2: the counts of insects found in agricultural experimental units treated with six different insecticides A-F

Chapter 3

Probability

3.1 Formal Definition of Probability

3.1.1 σ-algebra

Definition 3. $\mathcal{F}$, a collection of subsets of a set S, is called a ***σ-algebra*** associated with S if:

(a) $S \in \mathcal{F}$,

(b) $\mathcal{F}$ is closed under complements w.r.t. S:

$$E \in \mathcal{F} \Longrightarrow \overline{E} \in \mathcal{F},$$

(c) $\mathcal{F}$ is closed under countable unions:

$$E_1, E_2, \ldots \in \mathcal{F} \Longrightarrow \bigcup_{i=1}^{\infty} E_i \in \mathcal{F}.$$

Comments Definition 3 implies two facts.

1. $\mathcal{F}$ must contain the empty set $\emptyset$.

   Proof. Since $S \in \mathcal{F}$, we have $\overline{S} = \emptyset \in \mathcal{F}$. □

2. $\mathcal{F}$ must be closed under countable intersections.

   Proof. Let $E_1, E_2, \ldots \in \mathcal{F}$. We can then imply the following:

$$\overline{E_1}, \overline{E_2}, \ldots \in \mathcal{F} \Rightarrow \bigcup_i \overline{E_i} \in \mathcal{F} \Rightarrow \overline{\bigcup_i \overline{E_i}} \in \mathcal{F} \xrightarrow{\text{De Morgan's Law}} \bigcap_i E_i \in \mathcal{F}.$$

□

In short, we can take unions, intersections, and complements of members of $\mathcal{F}$ in any combination and the result will always be a member of $\mathcal{F}$.

3.1.2 Probability Measure

(Kolmogorov's axioms of probability) Definition 4. A ***probability measure*** P is a function $P : \mathcal{F} \mapsto \mathbb{R}$ satisfying

(a) $P(E) \geq 0 \; \forall E \in \mathcal{F}$,

(b) $P(S) = 1$,

(c) If $E_1, E_2, \ldots \in \mathcal{F}$ are disjoint (i.e. $E_i \cap E_j = \emptyset \; \forall i \neq j$) then

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i).$$

The triplet $(S, \mathcal{F}, P)$, consisting of a set S, a σ-algebra $\mathcal{F}$ of subsets of S, and a probability measure P, is called a ***probability space***.

Comments

- The ***sample space*** (S) is the set of all possible outcomes of an experiment.
- The ***event space*** $(\mathcal{F})$ is the set of possible events, where an ***event*** E is a subset of the sample space, $E \subseteq S$. An ***elementary event*** is one that consist of a single element of S, i.e. a singleton.
- The probability measure (P) has three important interpretations:

  1. **classical**: Different outcomes in the sample space S are "equally likely",
  2. **frequentist**: the relative frequency of an event over many trials,
  3. **subjective**: a numerical measure of the degree of belief held by an individual.

  Example 5. "A sensor can detect items within 10 cm of the sensor. The sensor is placed in a room together with an object, and the probability that the sensor makes a detection is 0.0001."

  1. **classical**: The volume within 10 cm of the sensor divided by the volume of the room is 0.0001.
  2. **frequentist**: If we repeat the experiment a lot of times, then the fraction of the experiments in which the sensor makes a detection is 0.0001.
  3. **subjective**: Someone's subjective degree of belief, measured on a numerical scale from 0 to 1, that the sensor will detect is 0.0001.

- several results that can be derived from the probability measure axioms:

  - $P(\emptyset) = 0$.

- $P(E) \leq 1$.
- $P(\overline{E}) = 1 - P(E)$.
- $P(E \cup F) = P(E) + P(F) - P(E \cap F)$..
- $P(E \cap \overline{F}) = P(E) - P(E \cap F)$.
- If $E \subset F$ then $P(E) \leq P(F)$.

3.2 Conditional Probability

Definition 6. If $P(F) > 0$ then the ***conditional probability*** of E given F is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Comments

- Difference among the following forms:
  - $P(E|F)$ – ***conditional probabilities***,
  - $P(E \cap F)$ – ***joint probabilities***,
  - $P(E)$ – ***marginal probabilities***.
- several results derived from the conditional probability definition:
  - $P(E|F) \geq 0$ for any event E.
  - $P(F|F) = 1$.
  - If the events $E_1, E_1, \ldots$ are pairwise disjoint, then $P\left(\left(\bigcup_i E_i\right)|F\right) = \sum_i P(E_i|F)$.
- Warning: In general, $P(E|F) \neq P(F|E)$.

Example 7. A medical test for a disease D has outcomes $+$ and $-$. The probabilities are

	D	$\overline{D}$	
$+$	0.009	0.099	0.108
$-$	0.001	0.891	0.892
	0.01	0.99	

By the definition of conditional probability, we have

$$P(+|D) = 90\%, \quad P(-|\overline{D}) = 90\%, \quad P(D|+) = \frac{0.009}{0.108} \approx 0.083.$$

The first two probabilities show that the test is fairly accurate. Sick people yield a positive 90% of the time and healthy people yield a negative 90% of the time.

3.3 Independence

Definition 8. Two events E and F are ***independent*** iff

$$P(E \cap F) = P(E)P(F).$$

Comments

- Extension: The events $E_1, \dots, E_k$ are independent if, for every subset of events of size $l \leq k$, say indexed by $\{i_1, \dots, i_l\}$,

$$P\left(\bigcap_{j=1}^{l} E_{i_j}\right) = \prod_{j=1}^{l} P(E_{i_j}).$$

- Independence could be either assumed or verified via the definition.
- Disjoint events with positive probability are not independent.
- From the definition of conditional probability, we can deduce that E and F are independent iff $P(E|F) = P(E)$.

Definition 9. For three events E_1, E_2, F, the pair of events E_1 and E_2 are said to be ***conditionally independent given*** F iff

$$P(E_1 \cap E_2|F) = P(E_1|F)P(E_2|F).$$

which could also be written as $E_1 \perp E_2 | F$.

3.4 Bayes' Theorem

(The Law of Total Probability) Theorem 10. Let $E_1, E_2, \dots$ be a partition of S, i.e. $E_i \cap E_j = \emptyset$ for $i \neq j$ and $\bigcup_i E_i = S$. Then, for any event $F \subseteq S$, we have

$$P(F) = \sum_i P(F|E_i)P(E_i).$$

Proof. $P(F) = P(\bigcup_i F \cap E_i) = \sum_i P(F \cap E_i) = \sum_i P(F|E_i)P(E_i)$. $\square$

(Bayes' Theorem) Theorem 11. If $P(F) > 0$ and let $E_1, E_2, \ldots$ be a partition on S s.t. $P(E_i) > 0 \, \forall i$, we have

$$P(E_i|F) = \frac{P(F|E_i)P(E_i)}{P(F)} = \frac{P(F|E_i)P(E_i)}{\sum_j P(F|E_j)P(E_j)},$$

where $P(E_i|F)$ is called the ***posterior***, $P(F|E_i)$ is called the ***likelihood***, $P(E_i)$ is called the ***prior***, and $P(F)$ is called the ***evidence***.

Proof. Exercise! haha □

Example 12. A new covid-19 test is claimed to correctly identify 95% of people who are really covid-positive and 98% of people who are really covid-negative. If only 1 in a 1000 of the population are infected, what is the probability that a randomly selected person who tests positive actually has the disease?

Let I = "has a covid infection" and T = "test is positive". We are given $P(T|I) = 0.95, P(\overline{T}|\overline{I}) = 0.98, P(I) = 0.001$. We can thus derive that

$$P(I|T) = \frac{P(T|I)P(I)}{P(T)} = \frac{P(T|I)P(I)}{P(T|I)P(I) + P(T|\overline{I})P(\overline{I})} = \frac{0.95 \times 0.001}{0.95 \times 0.001 + 0.02 \times 0.999} = 0.045.$$

Chapter 4

Discrete Random Variables

4.1 Random Variables

Definition 13. A ***random variable*** is a (measurable) mapping

$$X : S \mapsto \mathbb{R}$$

with the property that $\{s \in S : X(s) \leq x\} \in \mathcal{F} \ \forall x \in \mathbb{R}$. This ensures that any set $B \subseteq \mathbb{R}$ corresponds to an event in the event space $\mathcal{F}$.

Definition 14. The image of S under X is called the ***range*** of the random variable

$$\mathbb{X} \equiv X(S) = \{X(s) | s \in S\} = \{x \in \mathbb{R} \,|\, \exists s \in S \text{ s.t. } X(s) = x\}.$$

So S contains all the possible outcomes of the experiment, $\mathbb{X}$ contains all the possible outcomes of the random variable X.

Definition 15. The ***probability distribution*** of X is defined as

$$P_X = P_X(X \in B \subseteq \mathbb{R}) = P(\{s \in S : X(s) \in B\})$$

which enables us to transfer the probability measure P defined on $\mathcal{F}$ to the real numbers in a natural way, and vice versa. For instance,

$$P_X(X = 7) = P(\{s \in S | X(s) = 7\}),$$
$$P_X(a < X \leq b) = P(\{s \in S | a < X(s) \leq b\}).$$

Example 16. Consider counting the number of heads in a sequence of 3 coin tosses. The underlying sample space is

$$S = \{TTT, TTH, THT, HTT, THH, HTH, HHT, HHH\}.$$

Since we are only interested in the number of heads in each sequence, we define the random variable X by

$$X(s) = \begin{cases} 0, & s = TTT, \\ 1, & s \in \{TTH, THT, HTT\}, \\ 2, & s \in \{HHT, HTH, THH\}, \\ 3, & s = HHH. \end{cases}$$

Thus, the probability of the number of heads X is less than 2 is

$$\begin{aligned} P_X(X < 2) &= P(\{s \in S : X(s) < 2\}) \\ &= P(\{TTT, TTH, THT, HTT\}) \\ &= \frac{|\{TTT, TTH, THT, HTT\}|}{|S|} \\ &= \frac{4}{8} = \frac{1}{2}. \end{aligned}$$

On a side note, the above process uses the classical interpretation on the probability measure.

Definition 17. The ***Cumulative Distribution Function (CDF)*** of a random variable X is the function $F_X : \mathbb{R} \mapsto [0,1]$, defined by

$$F_X(x) = P_X(X \le x) = P(\{s \in S : X(s) \le x\}).$$

Comments

- Given a right-continuous function $F_X(x)$, check the following to verify if it is a valid CDF:
  - (i) $0 \le F_X(x) \le 1 \; \forall x \in \mathbb{R}$,
  - (ii) Monotonicity (non-decreasing): $\forall x_1, x_2 \in \mathbb{R}, x_1 < x_2 \Rightarrow F_X(x_1) \le F_X(x_2)$.
  - (iii) $F_X(-\infty) = 0, F_X(\infty) = 1$.
- For finite intervals $(a, b] \subseteq \mathbb{R}$, it is easy to check that

$$P_X(a < X \le B) = F_X(b) - F_X(a).$$

- Usually we suppress the subscript of $P_X(\cdot)$ and just write $P(\cdot)$ for the probability measure for the random variable, unless there is any ambiguity.

4.2 Discrete Random Variables

Definition 18. A random variable X is ***discrete*** if the range of X, $\mathbb{X}$, is countable, that is

$$\mathbb{X} = \{x_1, x_2, \ldots, x_n\} \text{ (finite)} \quad \text{or} \quad \mathbb{X} = \{x_1, x_2, \ldots\} \text{ (infinite)}.$$

Definition 19. For a discrete random variable X, we define the ***Probability Mass Function (PMF)*** as

$$p_X(x) = P_X(X = x), \quad x \in \mathbb{X}.$$

For completeness, we also define

$$p_X(x) = 0, \quad x \notin \mathbb{X}.$$

so that p_x is defined for all $x \in \mathbb{R}$.

Definition 20. The ***support*** of a random variable X is defined as

$$\left\{x \in \mathbb{R} : p_X(x) > 0\right\},$$

which is <u>almost always</u> the same as the range $\mathbb{X}$.

Properties of p_X and F_X

- $p_X(x_i) \geq 0$.
- $\sum_{x \in \mathbb{X}} p_X(x) = 1$.
- $F_X(x) = P(X \leq x)$, $x \in \mathbb{R}$.
- Let X be a discrete random variable with range $\mathbb{X} = \{x_1, x_2, \ldots\}$, where $x_1 < x_2 < \ldots$. Then for any $x \in \mathbb{R}$, if $x < x_1$, $F_X(x) = 0$; otherwise

$$F_X(x) = \sum_{x_i \leq x} p_X(x_i) \iff p_X(x_i) = F_X(x_i) - F_X(x_{i-1}), \quad i = 2, 3, \ldots,$$

  with $p_X(x_1) = F_X(x_1)$.
- $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to \infty} F_X(x) = 1$.
- F_X is continuous from the right on $\mathbb{R}$, i.e. for $x \in \mathbb{R}$, $\lim_{h \to 0^+} F_X(x + h) = F_X(x)$.
- F_X is non-decreasing, i.e. $a < b \Longrightarrow F_X(a) \leq F_X(b)$.
- For $a < b$, $P(a < X \leq b) = F_X(b) - F_X(a)$.

4.3 Functions of a Discrete Random Variable

Definition 21. The PMF of $Y = g(X)$ is found by grouping all the values in the range of x that correspond to the same value of Y, i.e.

$$p_Y(y) = \sum_{x \in \mathbb{X}: g(x)=y} p_X(x).$$

4.4 Mean and Variance

Definition 22. The ***expected value***, or ***mean*** of a discrete random variable X is defined to be

$$E_X(X) = \sum_{x \in \mathbb{X}} x p_X(x),$$

which is often written as $E(X)$, $E[X]$, or μ_X.

Theorem 23.

$$E(g(X)) = \sum_{x \in \mathbb{X}} g(x) p_X(x).$$

Proof. Let $Y = g(X)$, then

$$\begin{aligned} E(Y) &= \sum_{y \in \mathbb{Y}} y p_Y(y) \\ &= \sum_{y \in \mathbb{Y}} y \sum_{x \in \mathbb{X}: g(x)=y} p_X(x) \\ &= \sum_{y \in \mathbb{Y}} \sum_{x \in \mathbb{X}: g(x)=y} g(x) p_X(x) \\ &= \sum_{x \in \mathbb{X}} g(x) p_X(x). \end{aligned}$$

□

Theorem 24. Let X be a random variable with p_X. Let g and h be real-valued functions, $g, h : \mathbb{R} \mapsto \mathbb{R}$, and let a and b be constants. Then

$$E(ag(X) + bh(X)) = aE(g(X)) + bE(h(X)).$$

Proof. Exercise! □

Definition 25. Let X be a random variable. The ***variance*** of X, denoted by σ^2 or σ_X^2 or $\mathrm{Var}_X(X)$, is defined by

$$\mathrm{Var}_X(X) = E_X\left[(X - E_X(X))^2\right].$$

Proposition 26.

$$\mathrm{Var}(X) = E(X^2) - E(X)^2.$$

Proof.

$$\begin{aligned}\text{LHS} &= E\left[X^2 - 2E(X)X + E(X)^2\right] \\ &= E(X^2) - 2E(X)E(X) + E(X)^2 \\ &= \text{RHS}.\end{aligned}$$

□

Proposition 27.

$$\mathrm{Var}(aX \pm bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) \pm 2ab\mathrm{Cov}(X, Y).$$

Proof. Exercise! □

Definition 28. The ***standard deviation*** of a random variable X, written $\mathrm{sd}_X(X)$ or σ_X, is the square root of the variance,

$$\sigma_X = \sqrt{\mathrm{Var}_X(X)}.$$

Definition 29. The ***skewness*** (γ_1) of a discrete random variable X is given by

$$\gamma_1 = \frac{E_X\left[\left\{X - E_X(X)\right\}^3\right]}{\sigma_X^3}.$$

Sums of Random Variables

Let $X_1, X_2, \ldots, X_n$ be n random variables, perhaps with different distributions and not necessarily independent. Let $S_n = \sum_{i=1}^n X_i$ be the sum of those variables, and $\frac{S_n}{n}$ be their sample average. Both S_n and $\overline{S} = \frac{S_n}{n}$ are random variables themselves.

The mean of S_n and $\frac{S_n}{n}$ are given by

$$E(S_n) = \sum_{i=1}^n E(X_i), \quad E\left(\overline{S}\right) = \frac{\sum_{i=1}^n E(X_i)}{n} = \mu_X.$$

If $X_1, X_2, \ldots, X_n$ are **independent**, we can calculate the variance of S_n and $\overline{S} = \frac{S_n}{n}$ as well:

$$\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i), \quad \mathrm{Var}(\overline{S}) = \frac{\sum_{i=1}^{n} \mathrm{Var}(X_i)}{n^2} = \frac{\sigma_X^2}{n}.$$

4.5 Some important Discrete Random Variables

Definition 30. We say X follows a ***Bernoulli Distribution*** if $X \sim \mathrm{Bernoulli}(p)$, where $0 \leq p \leq 1$, and the pmf is given by

$$p_X(x) = \begin{cases} p & x = 1 \\ 1-p & x = 0 \\ 0 & \text{otherwise} \end{cases} = p^x (1-p)^{1-x}, \quad x \in \mathbb{X} = \{0, 1\}.$$

Definition 31. We say X follows a ***Binomial Distribution*** if $X \sim \mathrm{Binomial}(n, p)$, where $0 \leq p \leq 1$ and $n \in \mathbb{Z}^+$, and the pmf is given by

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \mathbb{X} = \{0, 1, 2, \ldots, n\}.$$

Definition 32. We say X follows a ***Geometric Distribution*** if $X \sim \mathrm{Geometric}(p)$, where $0 \leq p \leq 1$, and the pmf is given by

$$p_X(x) = p(1-p)^{x-1}, \quad x \in \mathbb{X} = \{1, 2, \ldots\}.$$

Alternatively, let $Y = X - 1$, then $Y \sim \mathrm{Geometric}(p)$ with the pmf

$$p_Y(y) = p(1-p)^y, \quad y \in \mathbb{N} = \{0, 1, 2, \ldots\}.$$

Definition 33. We say X follows a ***Poisson Distribution*** if $X \sim \mathrm{Poissons}(\lambda)$, where $\lambda > 0$, and the pmf is given by

$$p_X(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x \in \mathbb{X} = \{0, 1, 2, \ldots\}.$$

Definition 34. We say X follows a ***Discrete Uniform Distribution*** if $X \sim$ Uniform({1, 2, ..., n}), and the pmf is given by

$$p_X(x) = \frac{1}{n}, \quad x \in \mathbb{X} = \{1, 2, \ldots, n\}.$$

Table 4.1: Means and Variances of different distributions

	Mean(μ)	Variance(σ^2)	Skewness(γ_1)
Bernoulli	p	$p(1-p)$	N.A.
Binomial	np	$np(1-p)$	$\frac{1-2p}{\sqrt{np(1-p)}}$
Geometric(original)	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{2-p}{\sqrt{1-p}}$
Geometric(alternative)	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$	$\frac{2-p}{\sqrt{1-p}}$
Poisson	λ	λ	$\frac{1}{\sqrt{\lambda}}$
Uniform	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	0

Comments

- From table 4.1, we can see that the skewness of both Geometric and Poisson Distribution is always positive.
- **Approximation of Bionomial distribution as Poisson distribution**. It can be shown that for Binomial(n, p), when p is small and n is large, this distribution can be well approximated by the Poisson distribution with rate parameter $\lambda = np$, Poisson(np).